## REMARKS

### I. Prosecution History and Explanation of Amendments.

The application as filed contained 11 claims. In an official communication (Paper No. 10) dated October 3, 2002, claims 1-11 were subjected to a restriction requirement. In a responsive Amendment and Response to Restriction Requirement (Paper No. 13) filed on March 17, 2003, the Applicants added claims 12-13; and elected with traverse to prosecute the invention of Group I, claims 1, 2, 5, 6, 8, and 9, directed to a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO: 1. At the time of issuance of the outstanding Office Action, claims 1, 2, 5, 6, 8, 9, and 11-13 were pending in the application. In an Office Action dated June 24, 2003, claims 1, 2, 5, 6, 8, 9, and 11-13 were rejected variously under 35 U.S.C. §101, §102, §103, and §112, first and second paragraphs. Claims 1, 2, 5, 8, 9, and 11-13 have been amended and new claim 14 has been added.

### II. Explanation of amendments.

The amendments to the specification and the claims and the response made herein adopt suggestions made by the Examiner in an effort to advance this application toward allowance. The claim for priority finds support in the original filing papers of this application. Other amendments to the specification are intended solely to correct typographical errors. No new matter has been added.

Claims 1, 2, 5, 8, 9, and 11-13 as amended find support throughout the application (see especially Examples 1-8), and adopt the Examiner's recommendations to overcome the rejections. New Claim 14 finds support in the application at page 6, line 23 to page 7, line 1. The Applicants do not intend by these or any other amendments to abandon the subject matter of any claim as originally presented, and reserve the right to pursue such subject matter in other applications, such as continuing applications and divisional applications.

Fig. 7 as amended finds support throughout the application (see especially SEQ ID NOs: 1-2 and the amino acid alignment illustrated in Figs. 5-6).

## III. The Patent Office's rejection of claims 9 and 11 under 35 U.S.C. §101 should be withdrawn.

At page 3, paragraph 7 of the Office Action, the Patent Office rejected claims 9 and 11 under 35 U.S.C. §101, alleging that the claims were "not differentiable from the protein as it exists in nature." The Examiner suggested that the terms "isolated and purified" may be inserted into claim 9. Claims 9 and 11 as amended now recite an "isolated collectin protein" in accordance with the Patent Office's suggestion. Because the bases for rejection of claim 9 and its dependent claim 11 for relating to non-statutory subject matter is now moot, the rejection should be withdrawn. Likewise, this rejection should not be extended to new claim 14 which depends from claim 11.

## IV. The Patent Office's rejection of claim 12 under 35 U.S.C. §112, first paragraph should be withdrawn.

At page 4, paragraph 7 of the Office Action, the Patent Office rejected claim 1-12 under 35 U.S.C. § 112, first paragraph, alleging that these claims were not enabled by the specification. The Patent Office alleged that the claims should recite starting with the complement to the coding strand (SEQ ID NO: 1) because strands that hybridize to the coding strand would not encode a collectin.

Claim 12 as amended now recites starting with a " a probe which is complementary to the nucleotide sequence set out in SEQ ID NO: 1" in accordance with the Patent Office's suggestion. Because the basis for rejection of claim 12 for lack of enablement is now moot, the rejection should be withdrawn.

## V. The Patent Office's rejection of claims 5 and 11 under 35 U.S.C. §112, second paragraph, should be withdrawn.

At page 4, paragraph 6 of the Office Action, the Patent Office rejected claims 5 and 11 under 35 U.S.C. § 112, second paragraph, alleging that these claims were indefinite. As the basis for rejection, the Patent Office alleged that claim 5 was indefinite because:

> "the sequence of bases even where they hybridize to the noncoding complementary strand and as a sequence that encodes a protein with antiviral function, a Ca+2 carbohydrate recognition domain, a neck region, a collagen like region, and an N-terminal cysteine only define the function but not the

> sequence of the polynucleotide that was claimed by a structure. Where claim 5 recites SEQ ID NO:1, the size of the noncoding complementary strand is undefined. The claim should indicate the SEQ ID NO of the encoded protein as well. The blocking agent recited in line 5 does not appear to be defined as to what is the agent in the claim nor in the specification and the present response from applicant does not point to any particular pages and lines of the specification that define the blocking agent."

In addition, the Patent Office alleged that claim 11 was indefinite because neither claim 8 nor claim 9, from which claim 11 depends,

> "permit variation of deletion, substitution, and/or addition of one or more amino acids in the collectin protein. Claim 11 is unclear as to how deletions, substitutions, and/or additions which are not permissible from the language of current claims 8 and 9 from which claim 11 depends is narrowed by expanding the limits of claims 8 and 9 by the language of claim 11."

The rejection of claims 5 and 11 under 35 U.S.C. §112, second paragraph for indefiniteness should be withdrawn because claim 5 recites that the polynucleotide comprises a nucleotide sequence which hybridizes to a non-coding strand complementary to that set out in SEQ ID NO:1. Hybridization conditions are a common and accepted way to define a genus of polynucleotides. Claim 5 is consistent with Example 9 of the PTO's Training Materials for the Interim Written Description and Utility Guidelines. In Examples 1-4 of the specification the claimed polynucleotide is obtained by the steps comprising: 1) identifying molecules with highly conserved regions with the collectin set out in SEQ ID NO:3 (see Example 1), preparing probes from the clones so obtained for screening (see Example 2), screening by hybridization of a cDNA library from human liver for complementary polynucleotides (see Example 3), and determining the base sequences of the cDNA (see Example 4). It should be noted that the Applicant succeeded in Example 3 through screening-by-hybridization to isolate polynucleotides which hybridize with the probes prepared in Example 2. From this process, the specification teaches one of ordinary skill in the art how to isolate the polynucleotides of claim 5. Protein encoded by the polynucleotides of claim 5 have been show to have the same activities (e.g. anti-viral activity, immuno-enhancing activity) as known collectins. Claim 5 is directed to a polynucleotide that has the definite sequence properties to hybridize under the specified conditions and therefore the rejection should be withdrawn.

In addition, the term "blocking agent" is a term familiar to one of ordinary skill in the art of polynucleotide hybridization. Sambrook et al. Molecular Cloning: A Laboratory Manual ($2^{nd}$ ed., Cold Spring Harbor Laboratory: New York ,1989), p. 9.49, Table 9.1 lists a variety of blocking agents used that were used in hybridization experiments as far back as 1989. While not limited by the definition given in the specification, applicants point out that the term "blocking agent" recited in line 5 is definite because it can be defined, for example as DIG buffer II in Example 3 at page 10, line 30 of the specification. Therefore claims 5 and 11, which specify a blocking agent, are definite.

Finally, amended claim 11, depending from claims 8 and 9 no longer recites "deletion, substitution and/or addition of one or more amino acids." Amended claim 11 now specifies an isolated collectin protein according to Claim 8 or 9, wherein the collectin protein consists of (1) a $Ca^{2+}$-dependent carbohydrate recognition domain (CRD), (2) a neck region, (3) a collagen-like regions, and (4) an N-terminal region containing cysteine. The rejection should now be withdrawn with respect to claims 5 and 11 because the metes and bounds of the claims can be clearly determined. Likewise, this rejection should not be extended to new claim 14 which depends from claim 11.

## VI. The Patent Office's rejection of claims 1, 2, 6, 8, 9, and 11 under 35 U.S.C. §102(b) should be withdrawn.

To clarify the record, at page 6, paragraph 1 of the Office Action, the Patent Office alleges that: "This application currently names joint inventors." However, the patent application as originally filed indicated that Dr. Nobutaka Wakamiya was and is the sole inventor.

In addition, at page 6, paragraph 2 of the Office Action, the Patent Office rejected claims 8, 9, and 11 under 35 U.S.C. §102(b), alleging that the claims were anticipated by Hoppe et al.. In the Office Action, the Patent Office stated that:

> "The Hoppe et al. reference disclosed collectins which, absent factual evidence to the contrary, are anticipatory of claims 8, 9, and 11 because they contain any one or more deletions, substitutions, and/or additions . . . . Even where claims 8 and 9 recite a sequence, when the claims include via comprises and the deletion, substitution, and/or addition language, the actual sequence is not defined and therefore, the polypeptide can have

> any sequence such as would have been the polypeptide described in the Hoppe et al. reference." (Office Action at page 6.)

Applicants respectfully traverse.

The rejection based on Hoppe et al., should be withdrawn because claims 8, 9, and 11, as amended, no longer recite "comprising" or "deletion, substitution and/or addition of one or more amino acids." Amended claim 8 now specifies an isolated collectin protein consisting of the amino acid sequence set out in SEQ ID NO:2. Amended claim 9 now specifies an isolated collectin protein consisting of the amino acid sequence encoded by the nucleotide sequence set out in SEQ ID NO: 1. As noted above, amended claim 11 now specifies an isolated collectin protein according to Claim 8 or 9, wherein the collectin protein consists of (1) a $Ca^{2+}$-dependent carbohydrate recognition domain (CRD), (2) a neck region, (3) a collagen-like regions, and (4) an N-terminal region containing cysteine.

The rejection should now be withdrawn with respect to claims 8, 9, and 11 because Hoppe et al. does not teach an isolated collectin protein consisting of the amino acid sequence set out in SEQ ID NO:2 (claims 8 or 11) or an isolated collectin protein consisting of the amino acid sequence encoded by the nucleotide sequence set out in SEQ ID NO: 1 (claims 9 and 11). Likewise, this rejection should not be extended to new claim 14 which depends from claim 11. New claim 14 is duly supported by the amino acid sequences set out in SEQ ID NOS: 1 and 2 and the amino acid alignment illustrated in Figures 5 and 6.

In addition, at page 7, paragraph 1 of the Office Action, the Patent Office rejected claims 5 and 13 under 35 U.S.C. §102(b), alleging that these claims were also anticipated by Hoppe et al. In the Office Action, the Patent Office stated that the Hoppe et al. reference disclosed collectins which, "absent factual evidence to the contrary, are anticipatory of claims 5 and 13 because claim 5 contains no sequence recitation as the result of the hybridization parameters." (Office Action at page 6.) Applicants respectfully traverse.

The rejection based on Hoppe et al. should be withdrawn because amended claim 5 and claim 13, which depends thereon, require that the sequence hybridize to SEQ ID NO: 1 under certain parameters and because the nucleotide sequence disclosed in Hoppe et al. would not hybridize under the conditions recited. It is well known in the art that the hybridization temperature is lowered by about 1.4°C for every 1% mismatch in the

hybridizing DNAs. (*See* DNA - Basics of Structure and Analysis available at http://www.ndsu.nodak.edu/instruct/mcclean/;lsc731/dna/dna6.htm (Oct. 8, 2003). The DNA sequence disclosed in Hoppe et al. that is closest to the sequence as claimed is the Human Mannan Binding Protein (MBP1) cited in Table 1 (Taylor et al. attached hereto as Appendix 2). There is no significant similarity (*See* Blast 2 Sequences results attached hereto as Appendix 3) between SEQ ID NO: 1 in the claims and the DNA sequence of human MBP1 (Taylor et al., 1989), cited as the 102(b) art in Hoppe et al. by the examiner. Because there is no significant similarity of amino acid sequences (31% identity), there would be no significant similarity of the nucleotide sequences, and there would be no hybridization of the two DNA sequences, under the hybridization conditions cited in the claims. SEQ ID NO: 1 in the instant application is 44 % GC (704/1595 bases) So in 5X SSC (($\log M [Na^+]$) = log 0.825) its approximate melting temperature would be: Eff $T_m = 81.5 + 16.6(\log M [Na^+]) + 0.41(\%G+C)$. Therefore, the Eff $T_m = 81.5 + 16.6 (\log 0.825) + 0.41(44)°C = 98.2°C$. Thus, in a hybridization that is performed at 55°C (43.2°C lower than the melting temperature; 100% - (43.2°C/1.4°C) = 69.2% homology) there would be no detectable hybridization of the 31% homologous MBP1 sequence because 31% is so far below the approximate melting temperature. Consequently, the rejection should now be withdrawn with respect to claims 5 and 13.

Further, at page 6, paragraph 5 of the Office Action, the Patent Office rejected claims 1, 2, and 6 under 35 U.S.C. §102(b), alleging that the claims were anticipated by GenBank Accession Number AB002631. In the Office Action, the Patent Office stated that:

> "While the reference would appear to be 24 Jan 1999, the record indicates a submission of 4 Apr 1997 which, absent factual evidence to the contrary, predates the instant application filing date and the foreign priority date." (Office Action at page 6.)

Applicants respectfully traverse.

Submitted herewith as Appendix 4 is the Rule 132 Declaration of Nobutaka Wakamiya in which he relates that the GenBank reference was not laid open to the public until June 24, 1999. In addition, submitted herewith as Appendix 5 is an email from Yoshio Tateno, Ph.D at the Center for Information Biology and DNA Data Bank of Japan, the collaborator with GenBank to which the sequence was submitted. As Dr. Tateno relates the

sequence data with accession number AB002631 was not laid open to the public until June 24, 1999. As a consequence, GenBank Accession Number AB002631 is not 102(b) prior art. Accordingly, the rejection on the basis of GenBank Accession Number AB002631 should be withdrawn.

Finally, at page 7, paragraph 5 of the Office Action, the Patent Office rejected claims 5, 12, and 13 under 35 U.S.C. §102(b), alleging that the claims were anticipated by Kawai et al.. In the Office Action, the Patent Office stated that:

> "claim 5 is anticipated since the claim does not per se require the polynucleotide obtained to must have the sequence of SEQ ID NO: 1 but only that it had hybridized under the recited conditions." (Office Action at page 8.)

Applicants respectfully traverse.

The rejection based on Kawai et al., should be withdrawn because amended claims 12 and 5, and claim 13 which depends thereon, now require that the sequence hybridize to SEQ ID NO: 1 under certain parameters and because the nucleotide sequence disclosed in Kawai et al. would not hybridize under the conditions recited. As mentioned above, it is well known in the art that the hybridization temperature is lowered by about 1.4°C for every 1% mismatch in the hybridizing DNAs. The DNA sequence disclosed in Kawai et al. that is closest to the sequence as claimed is the Bovine Mannan Binding Protein (MBP). There is no significant similarity (*See* Blast 2 Sequences results attached hereto as Appendix 6) between SEQ ID NO: 1 in the claims and the DNA sequence of Bovine MBP, cited as the 102(b) art in Kawai et al. by the examiner. Because there is no significant similarity of the amino acid sequences (33% identity), there would be no significant similarity of the nucleotide sequences and there would be no hybridization of the two DNA sequences, especially under the hybridization conditions cited in the claims. SEQ ID NO: 1 in the instant application is 44 % GC (704/1595 bases). So in 5X SSC ((log M [$Na^+$]) = log 0.825) its melting temperature would be: Eff $T_m = 81.5 + 16.6(\log M [Na^+]) + 0.41(\%G+C)$. Therefore, the Eff Tm = 81.5 + 16.6 (log 0.825) + 0.41(44)°C = 98.2°C. Thus, in a hybridization that is performed at 55°C (43.2°C lower than the melting temperature; 100% - (43.2°C/1.4°C) = 69.2% homology) there would be no detectable hybridization of the 33% homologous MBP1 sequence because 31% is so far below the approximate melting temperature. Consequently, the rejection should now be withdrawn with respect to claims 5, 12, and 13.

**VII. The Patent Office's rejection of claims 1, 2, 5, 6, 8, 9, and 11-13 under 35 U.S.C. §103 should be withdrawn.**

At page 8, paragraph 3 of the Office Action, the Patent Office rejected claims 1, 2, 5, 6, 8, 9, and 11-13 under 35 U.S.C. §103(a), alleging that the claims were obvious over Kawai et al. in view of GenBank Accession Number AB0022631 cited as a secondary reference. Applicants respectfully traverse.

The rejection based on the combination of Kawai et al. and GenBank Accession Number AB0022631, should be withdrawn because, as noted above, the GenBank Accession Number reference is not prior art on the basis of publication on the same day as applicant's filing date.

**CONCLUSION**

Withdrawal of the rejections and allowance of all pending claims in the application are respectfully requested in view of the foregoing amendments and remarks. Should the Examiner wish to discuss any issues of form or substance in order to expedite allowance of the pending application, she is invited to contact the undersigned agent at the number indicated below.

If there are any additional fees due in connection with the filing of this response, please charge the fees to our Deposit Account No. 13-2855. If a fee is required for an extension of time under 37 C.F.R. §1.136 not accounted for above, such an extension is requested and the fee should also be charged to our Deposit Account.

Dated: November 20, 2003

Respectfully submitted,

By [signature]

Mark H. Hopkins, Ph.D.
Registration No.: 44,775
MARSHALL, GERSTEIN & BORUN LLP
233 S. Wacker Drive, Suite 6300
Sears Tower
Chicago, Illinois 60606-6357
(312) 474-6300
Agent for Applicant

Biochem. J. (1989) 262, 763–771 (Printed in Great Britain)



# Structure and evolutionary origin of the gene encoding a human serum mannose-binding protein

Maureen E. TAYLOR,*‡ Paul M. BRICKELL,† Roger K. CRAIG†§ and John A. SUMMERFIELD*||
*Department of Medicine, St. Mary's Hospital Medical School, Norfolk Place, London W2 1NY,
and †The Medical Molecular Biology Unit, Biochemistry Department, University College and Middlesex School of Medicine, London W1P 6DB, U.K.

The *N*-terminal sequence of the major human serum mannose-binding protein (MBP1) was shown to be identical at all positions determined with the amino acid sequence predicted from a cDNA clone of a human liver MBP mRNA. An oligonucleotide corresponding to part of the sequence of this cDNA clone was used to isolate a cosmid genomic clone containing a homologous gene. The intron/exon structure of this gene was found to closely resemble that of the gene encoding a rat liver MBP (MBP A). The nucleotide sequence of the exons differed in several places from that of the human cDNA clone published by Ezekowitz, Day & Herman [(1988) J. Exp. Med. 167, 1034–1046]. The MBP molecule comprises a signal peptide, a cysteine-rich domain, a collagen-like domain, a 'neck' region and a carbohydrate-binding domain. Each domain is encoded by a separate exon. This genomic organization lends support to the hypothesis that the gene arose during evolution by a process of exon shuffling. Several consensus sequences that may be involved in controlling the expression of human serum MBP have been identified in the promoter region of the gene. The consensus sequences are consistent with the suggestion that this mammalian serum lectin is regulated as an acute-phase protein synthesized by the liver.

## INTRODUCTION

Proteins capable of binding mannose-terminated glycoproteins have been isolated from the liver and serum of rats [1–3], rabbits [4,5] and humans [6–9]. At least two mannose-binding proteins (MBPs) are present in human serum [8]. The major human serum MBP (MBP1) is a homopolymer of approx. 700 kDa composed of subunits of 32 kDa and is secreted by the liver [7]. MBP1 has specificities for *N*-acetylglucosamine, *N*-acetylmannosamine, glucose and fucose as well as mannose [8]. The other MBP (MBP2) is a homopolymer of approx. 200 kDa composed of subunits of 28 kDa and is specific only for mannose and fucose [8]. Both lectins are $Ca^{2+}$-dependent. The function of these MBPs is unknown, but they have a high affinity for yeast mannan [1–9] and thus might act as a part of the immune system, binding and inactivating yeasts and other micro-organisms bearing mannose oligosaccharides. Support for this hypothesis comes from the finding that serum MBPs from rat, rabbit and human can activate complement through the classical pathway [10].

Cloning and sequencing of two rat liver MBPs has shown that they are divided into several domains: a signal peptide, a short cysteine-rich *N*-terminal domain, a collagen-like domain, a 'neck' region and a *C*-terminal carbohydrate-recognition domain [11]. The primary structure of a human liver MBP has recently been determined and found to be similar to that of the rat MBPs [12]. Northern-blot hybridization revealed marked MBP mRNA accumulation in the liver of a trauma victim, suggesting that MBP may be an acute-phase reactant [12]. The sequences of carbohydrate-recognition domains of MBPs are highly homologous to those found in other carbohydrate-binding proteins, including membrane-bound lectins such as the rat asialoglycoprotein receptor [13]. By using the conserved amino acid residues found in the carbohydrate-recognition domains of these proteins as a guide, it has been possible to identify putative carbohydrate-recognition domains in proteins such as pulmonary-surfactant apoprotein [14] and the core protein of cartilage proteoglycan [15].

The gene structure of one rat liver MBP (MBP A) has been determined and found to share structural features with the genes for non-fibrillar collagens as well as with the genes for other carbohydrate-binding proteins [16]. In the present paper we give the structure of the gene encoding a human MBP, which we suggest is the gene encoding the major human serum MBP (MBP1). We show that the structure of this gene is very similar to that of the gene encoding rat liver MBP A. We have also identified consensus sequences in the promoter region of the human MBP gene that indicate that its expression is regulated as part of the acute-phase response.

## MATERIALS AND METHODS

### Protein sequence analysis

The major human serum MBP (MBP1) was isolated

Abbreviation used: MBP, mannose-binding protein.
‡ Present address: Department of Biochemistry, Columbia University, 630 West 168th Street, New York, NY 10032-3702, U.S.A.
§ Present address: Research Division 1, ICI Pharmaceuticals, Alderley Park, Cheshire SK10 4TG, U.K.
|| To whom correspondence should be addressed.
These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers X15954, X15955, X15956 and X15957.

from pooled serum collected from patients undergoing venesection for haemochromatosis by methods previously described [8]. Affinity chromatography with a monoclonal antibody against human serum MBP1 (Mab37) was used to purify MBP1 to homogeneity, as judged by SDS/polyacrylamide-gel electrophoresis. The *N*-terminal amino acid sequence of this serum MBP sample was determined on an Applied Biosystems gas-phase sequencer by Dr. K. Drickamer (Columbia University, New York, NY, U.S.A.).

### Isolation of clones

A cosmid library of human genomic DNA was kindly provided by Dr. Dimitris Kioussis (National Institute for Medical Research, London, U.K.). This library was constructed in the cosmid vector cos 202 by using partially digested *Sau*3AI fragments of genomic DNA from the human acute-lymphocytic-leukaemia cell line HPB-ALL [17]. Some 500000 colonies were screened for the presence of MBP-encoding inserts by the method of Grunstein & Hogness [18]. The probe used (probe 4) was a 40-base oligonucleotide with the sequence 5′-dGGGAACTCA-CAGACGGCAGATGGGAGGTGGAAGCAAGGG-G-3′ derived from the region of the cDNA clone [12] coding for the carbohydrate-binding domain of MBP (Fig. 3*d*, exon 4, bases 645–607, lower strand). The oligonucleotide was labelled with [$\gamma$-$^{32}$P]ATP to a specific radioactivity of $10^9$ c.p.m./$\mu$g by the use of T4 polynucleotide kinase [19]. Duplicate filters were prehybridized overnight at 55 °C in a solution composed of 6 × SSC, 5 × Denhardt's solution, 0.5% SDS and 100 $\mu$g of denatured herring testis DNA/ml and then hybridized for 8 h at 55 °C in the same solution containing the labelled probe at $10^6$ c.p.m./ml. [1 × SSC is 0.15 M-NaCl/15 mM-sodium citrate buffer, pH 7.0, and 1 × Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin (Pentax fraction V).] The filters were washed at 65 °C for 30 min in 2 × SSC/0.1% SDS and autoradiographed. Colonies corresponding to positive signals were purified by two rounds of screening at low density.

### Analysis of clones

Four further oligonucleotide probes were used, each having a sequence found in the cDNA clone [12]. Probe 1 was 31 bases long with the sequence 5′-dGGTG-TGAGAAAACTCAGGGAAGGTTAATCTC-3′ derived from the 5′ untranslated region of the cDNA clone (Fig. 3*a*, exon 1, bases 57–27, lower strand). Probe 2 was 31 bases long with the sequence 5′-dGGCCCTTTG-GTCCTGGTGACCCAGAAGGCCC-3′ derived from the region of the cDNA clone encoding the collagen-like domain (Fig. 3*b*, exon 2, bases 580–550, lower strand). Probe 3 was 29 bases long with the sequence 5′-dCC-ATTCCTGGGGGTGGCCACAGAGGCTGG-3′ derived from the region of the cDNA clone encoding the carbohydrate-binding domain (Fig. 3*d*, exon 4, bases 408–379, lower strand). Probe 5 was 28 bases long with the sequence 5′-dGGTCATGTCATTAGGGGAGCC-CTCTACC-3′ derived from the 3′ untranslated region of the cDNA clone (Fig. 3*d*, exon 4, bases 1065–1038, lower strand). The sequences of some probes differed slightly from the sequence in Fig. 3 since they were based on the reported sequence of the cDNA clone [12], which we believe contains errors (see below). Southern-blot analysis was used to identify the restriction fragments of the cosmid insert that hybridized to the probes. Fragments containing sequence of interest were cloned into M13 mp18 and M13 mp19 and sequenced by the dideoxy method [20] with the use of Sequenase (United States Biochemical Corporation). The strategy for the nucleotide sequence analysis is shown in Fig. 2.

## RESULTS AND DISCUSSION

### Protein sequences of human liver and serum MBPs

Our aim was to isolate the gene encoding the major human serum MBP (MBP1) by using sequence data from the human liver MBP cDNA clone sequenced by Ezekowitz *et al.* [12]. However, since both human and rat liver contain at least two MBPs [9,11], we needed to show that the human liver cDNA sequenced by Ezekowitz *et al.* [12] encoded human serum MBP. The *N*-terminal amino acid sequence of serum MBP1 is shown in Fig. 1. At all 21 positions determined, this sequence matched perfectly the amino acid sequence predicted from the nucleotide sequence of the human liver MBP cDNA clone reported by Ezekowitz *et al.* [12]. This strongly suggests but does not prove absolutely that the cDNA clone encodes the major human serum MBP (MBP1).

### Isolation of positive clones

Screening of 500000 colonies with radiolabelled oligonucleotide probe 4 yielded one positive clone (cMBP1). This clone contained an insert of approx. 30 kb. Southern-blot analysis with oligonucleotide probes corresponding to sequences at the 5′ and 3′ ends of the cDNA clone (probes 1 and 4) showed that cMBP1 contained the whole of the transcribed region of the gene together with approx. 9 kb of 5′ flanking sequence and 16 kb of 3′ flanking sequence. Therefore no further screening was performed. A partial restriction map of the clone is shown in Fig. 2.

### Intron/exon structure of the gene

The nucleotide sequence of the 5′ flanking sequence, the exons and the intron/exon boundaries of the MBP gene contained within cMBP1 is shown in Fig. 3. Comparison of this nucleotide sequence with that of the cDNA clone [12] (Fig. 4) shows that the protein-coding region of the gene is interrupted by three introns of approx. 600 bp, 1350 bp and 800 bp in size. The sequences at all of the intron/exon boundaries contain the usual splice donor and acceptor consensus sequences [21]. The predicted amino acid sequence of the protein encoded by this gene is shown in Fig. 5.

The intron/exon structure and the sizes of the introns are almost identical with those of the rat MBP A gene [16]. Thus exon 1 encodes the signal peptide, a cysteine-rich domain and seven copies of the Gly-Xaa-Yaa motif typical of the triple-helix-forming region of collagen. At the site of the first intron the regular Gly-Xaa-Yaa pattern is interrupted by the sequence Gly-Gln-Gly. Exon 2 then encodes a further 12 Gly-Xaa-Yaa repeats. Both rat MBP A and rat MBP C have the same single Gly-Gln-Gly interruption in the Gly-Xaa-Yaa repeat structure. A regular Gly-Xaa-Yaa pattern encoded by two adjacent exons is typical of genes encoding other non-fibrillar collagens [22–25]. Exon 3 encodes a 'neck' region and exon 4 encodes a carbohydrate-recognition domain. This is also the case for the rat MBP A gene [16]. Moreover, the carbohydrate-binding domain of the rat

ND-Thr-Val-Thr-ND-Glu-ND-Ala-Gln-Lys-Thr-ND-Pro-Ala-ND-Ile-ND-ND-ND-ND-Pro-Gly-Ile-Asn-Gly-Phe-Pro-Gly-Lys-Asp

**Fig. 1. *N*-Terminal amino acid sequence of the major human serum mannose-binding protein (MBP1)**

Abbreviation: ND, not determined.

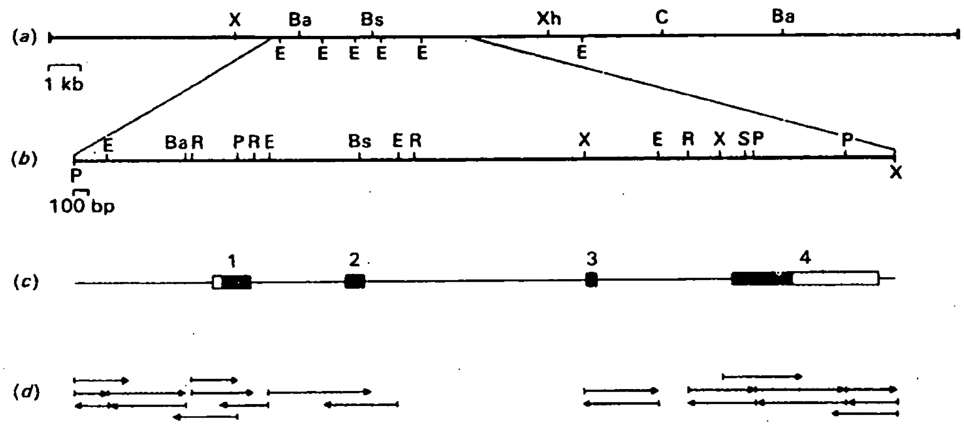



**Fig. 2. Partial restriction map of cosmid genomic clone cMBP1 containing the gene encoding human MBP1**

(*a*) Insert of cosmid clone cMBP1. (*b*) Region of cMBP1 sequenced. (*c*) Structure of MBP1 gene. (*d*) Sequencing strategy. Abbreviations: Ba, *Bam*HI; Bs, *Bst*EII; C, *Cla*I; E, *Eco*RI; P, *Pst*I; R, *Rsa*I; S, *Stu*I; X, *Xba*I; Xh, *Xho*I. Coding regions are indicated by black boxes. 5′ and 3′ untranslated regions are indicated by open boxes.

liver asialoglycoprotein receptor is also encoded by discrete exons separated from the exons encoding other functional domains [26]. This is consistent with the hypothesis that evolutionary shuffling has allowed carbohydrate-recognition domains to become associated with different effector molecules [13].

The sequence of the exons was found to be different from the published cDNA sequence [12] in several places, as shown in Fig. 4. The 3′ untranslated region of the published cDNA clone [12] contains two stretches of 23 and 56 nucleotide residues that are absent from the genomic clone. In addition, nucleotide residues 975–1128, which are present in exon 4 (Fig. 3*d*, nucleotide residues 728–958) are absent from the 3′ untranslated region of the cDNA clone. It is unlikely that this reflects the presence of an intron in the 3′ untranslated region of the human MBP mRNA, since there are no consensus splice donor and acceptor sites at the appropriate positions in the genomic sequence (Fig. 3*d*). It is possible that these differences arose from reverse transcriptase errors in the construction of the cDNA clone or that they arose because the cDNA and genomic clones derive from two distinct genes. The remainder of the sequence differences are changes, deletions or insertions of single base-pairs that, in some cases, yield differences in the amino acid sequences predicted from the two nucleotide sequences.

The amino acid sequences predicted from our sequence, from the published cDNA sequence [12] and from the two rat MBPs [11] are shown in Fig. 5. As indicated, the signal sequence predicted from the genomic clone has one extra amino acid residue (residue 7) and four amino acid residue changes when compared with that predicted from the cDNA clone. The signal sequence predicted from the genomic clone retains the typical hydrophobic structure expected in a secreted eukaryotic protein [27] and, in particular, has a serine residue at position 20, immediately preceding the *N*-terminal glutamic acid residue of the mature protein (Fig. 5). This conforms with the general observation that the amino acid residue immediately preceding the site of cleavage by the signal peptidase has a small side chain [28]. There are three amino acid residue differences in the first part of the collagen-like domain in exon 1, and in exon 3 only six of the 22 amino acid residues are the same as those published. Finally, there are 16 amino acid residue differences in the carbohydrate-binding domain (exon 4). The amino acid residue changes increase the overall homology of the human MBP to the rat MBPs [11], especially in the carbohydrate-recognition domain. Thus in 21 positions a residue is changed to that found in rat MBP C and in 16 positions a residue is changed to that found in rat MBP A. In particular, Thr-187 predicted by cMBP1 is one of the 18 invariant residues found in all mammalian carbohydrate-binding proteins [29] whereas Pro-187 predicted by the cDNA clone is not. These results give values of 61 % for the homology of the human MBP to rat MBP C and of 52 % for the homology of the human MBP to rat MBP A.

It is possible the sequence of the cDNA clone [12] differed from that of our genomic clone because the genomic clone derives from a second MBP gene. However, there are two lines of evidence that suggest that the differences arise because the sequence of the cDNA clone [12] was partially incorrect. Firstly, where there are stretches of amino acid residue differences (e.g. Fig. 5, residues 107–123) this reflects an alteration of the reading frame caused by a single-base deletion/insertion rather than extensive nucleotide differences (Fig. 4, nucleotides 389–440). Secondly, if the two clones were derived from different genes encoding closely related proteins we would

expect to find extensive nucleotide differences in the third-base position of codons, even when the predicted amino acid sequences were identical. This is seen when the rat and human nucleotide sequences are compared, but not when the human cDNA and genomic sequences are compared. For these reasons we suggest that the human cDNA [12] and genomic clones are derived from a single human MBP gene. In view of the identity of their products' predicted *N*-terminal sequence with the *N*-terminal sequence of the major human serum MBP (MBP1) (Fig. 1), we suggest that they derive from the human serum MBP1 gene. However, we cannot exclude the possibility that they represent two different MBP genes, in which case our data do not allow us to identify which clone, if either, represents the human serum MBP1 gene. If the cDNA [12] and genomic clones are derived

(*a*) 5′ flanking sequence, exon 1 and portion of intron 1

```
ctgcagtggagactgtctttgttttcaaagggaaacttggaggcttagacctatggggct -784
aggctgctgaggtttcttagggggcaatagctggaagaaagctctgaAGAACAAATGaaa -724
ggttaatactgagaaatgggaggaggattcaaggcaagttttctaattgccagtggtttt -664
tgactcacAGAACATggggaattcctgccagaaagtagagaggtatttagcactctgcca -604
gggccaacgtagTAAGAAATTTCCAGagaaaatgcttacccaggcaagcctgtctaaaac -544
accaaggggaagcaaactccagttaattctgggctgggttggtgactaaggttgaggttg -484
atctgaggttgagaccttcctctttggatcaccagctttcagctcagggcctgccaatga -424
gtaaatgatagttaacaggtcctggaggggaatcagctgcccagatacaaagatgggatt -364
caggtggcagatggacccgaagaggacatggagagaaagaggaagctcctacagacacct -304
gggtttccactcattctcattccctaagctaacaggcataagccagctggcaatgcacGG -244
TCCCATTTGTTCTcactgccaccgaaagcatgtttatagtCTTCCAGCAGCAACGCCAGG -184
tgtctaggcacagatgaacccctccttaggatccccactgctcatcatagtgcctacctt -124
tgttaaagtactagtcacgcagtgtcacaaggaatgtttacttttCCAAATccccagcta -64
gaggccagggatgggtcatctatttcTATATAgcctgcacccagattgtaggacagaggg -4
catGCTCGGTAAATATGTGTTCATTAACTGAGATTAACCTTCCCTGAGTTTTCTCACACC 57
AAGGTGAGGACCATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCA 117
            MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAla

GCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCCAAAAGACCTGCCCTGCAGTGATT 177
AlaSerTyrSerGluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIle

GCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGCGTGATGGCACCAAG 237
AlaCysSerSerProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLys

GGAGAAAAGGGGGAACCAGgtacgtgttgg                               267
GlyGluLysGlyGluProG
```

(*b*) Exon 2 and flanking sequences

```
aacaacaactactagcaaaacaaatgcagttaattttcactttgcaccctccctgcagca 60
acctccacgtggcaactttatttcttaagttattgctctcaggtgcacaccatacagtta 120
ttgagagcagtgctcagaaaggtcagtcctgggtcaaggtctcctttctcctgagaagga 180
ttggcatcaaactcttgaagagagagcaagaacatgagatattaagtcacattcctttgt 240
cttccaacagGCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCCA 300
          lyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProPro

GGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAA 360
GlyAsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLys

AGTCCGGgtaaggacccc                                           378
SerProA
```

(*c*) Exon 3 and flanking sequences

```
tctagATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGC 60
     spGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAl

ACGTATCAAAAAGTgtaagcttttctcttactctccaggcagcttgaagtttgggaaaa 120
aArgIleLysLysT

atagatgcaacaaatatttgttgaatgcatataattttctgtaccctgctaggcatttct 180

catattcttacctcatgaaattctcacaacattttggtagaaatggaggcaaagggaagt 240

taaattacttgttcaaatgcacagagctaataaatggcaggggtggtttatagatggaag 300

tcagtctgactcgagagaccctaatcctttaccgtctgatattgctcactgaaaatggga 360

cttgagatttgggtggtcactggtattgagacctggccgtggggtctaacctgcctgggg 420

caaatattttcaga 434
```

(*d*) Exon 4 and flanking sequences

```
accagtctgtctgttcacatttgggttgccatatttaacaataaaaagacaaagcacca 60

gttaaattggatttcagctgaaaaacattttagcacaagtgtgtctgaaaaattatatg 120

ggagatatactaaaatattcattgtttgtctgaaattcaaatttaactggacatactata 180

ttttatccggcaactctactctagaagactttttcttgagaaataccttgagttgggctt 240

aaggatgagtgagtttcacccacttttttcacattttagGGCTGACCTTCTCTCTGGGCAA 300
                                        rpLeuThrPheSerLeuGlyLy

ACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAA 360
sGlnValGlyAsnLysPhePheLeuThrAsnGlyGluIleMetThrPheGluLysValLy

GGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATGCTGCAGAGAATGG 420
sAlaLeuCysValLysPheGlnAlaSerValAlaThrProArgAsnAlaAlaGluAsnGl

AGCCATTCAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGA 480
yAlaIleGlnAsnLeuIleLysGluGluAlaPheLeuGlyIleThrAspGluLysThrGl

AGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAAACTGGAACGAGGGTGA 540
uGlyGlnPheValAspLeuThrGlyAsnArgLeuThrTyrThrAsnTrpAsnGluGlyGl

ACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAA 600
uProAsnAsnAlaGlySerAspGluAspCysValLeuLeuLeuLysAsnGlyGlnTrpAs

TGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCTGAAGGGTCAT 660
nAspValProCysSerThrSerHisLeuAlaValCysGluPheProIleEnd

ATCACTCAGGCCCTCCTTGTCTTTTTACTGCAACCCACAGGCCCACAGTATGCTTGAAAA 720

GATAAATTATATCAATTTCCTCATATCCAGCATTGTTCCTTTTGTGGGCAATCACTAAAA 780

ATGATCACTAACAGCACCAACAAAGCAATAATAGTAGTAGTAGTAGTTAGCAGCAGCAGT 840

AGTAGTCATGCTAATTATATAATATTTTTAATATATACTATGAGGCCCTATCTTTTGCAT 900

CCTACAATTAATTATCTAGTTTAATTAATCTGTAATGCTTTCGATAGTGTTAACTTGCTG 960

CAGTATGAAAATAAGACGGATTTATTTTTCCATTTACAACAAACACCTGTGCTCTGTTGA 1020

GCCTTCCTTTCTGTTTGGGTAGAGGGCTCCCCTAATGACATCACCACAGTTTAATACCAC 1080

AGCTTTTTACCAAGTTTCAGGTATTAAGAAAATCTATTTTTGTAACTTTCTCTATGAACT 1140

CTGTTTTCTTTCTAATGAGATATTAAACCATGTAaagaacataaataacaaatctcaagc 1200

aaacagcttcacaaattctcacacacatacatacctatatactcactttctaga 1254
```

**Fig. 3. Partial nucleotide sequence of human MBP1 gene**

Sequences from introns and the 5′ flanking region are shown in lower case. The consensus sequences in the 5′ flanking region are shown in upper case and are underlined (see the text for details). The four exons are shown in upper case. The sequences from which the five oligonucleotide probes were derived are marked with a line above them.

```
                                                                           77
GCTCGGTAAATATGTGTTCATTAACTGAGATTAACCTTCCCTGAGTTTTCTCACACCAAGGTGAGGACCATGTCCCT
GCTCGGTAAATATGT-TTCATTAACTGAGATTAACCTTCCCTGAGTTTTCTCACACCAAGG-GAG-ACCATGTCC-T
                                                                          154
GTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCAGCGTCTTACTCAGAAACTGTGACCTGTGAGGATGCCC
GTTTC-ATCACTCCCTCTC-TTCTCCTGAGTATGGTGGCAGCGTCTTAC-CAGAAACTGTGACCTGTGAGGATGCCC
                                                                          231
AAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGC
AAAAGACCTGCCCTGCAGTGATTGCCTGTAGCTCTCCAGGCATCAACGGCTTCCCAGGCAAAGATGGGCGTGATGGC
                        1  2                                              306
ACCAAGGG-AGAAAAGGGGGAACCAG|G-GCCAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCC
ACCAAGGGTAGAAAAGGGGGAAC-AG|GT--CAAGGGCTCAGAGGCTTACAGGGCCCCCCTGGAAAGTTGGGGCCTCC
                                                                 2  3     383
AGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGG|ATGGTGATA
AGGAAATCCAGGGCCTTCTGGGTCACCAGGACCAAAGGGCCAAAAAGGAGACCCTGGAAAAAGTCCGG|ATGGTGATA
                                                            3  4          457
GTAGCC-T-GGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAAAGT|GT-AAGCTGACCTTC
GTAGCCC-AGGCTGCCTCAGAAAGAAAAGCTCTGCAAACAGAAATGGCACGTATCAAAA-GT|G-C---CTGACCTTC
                                                                          534
TCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTT
TCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAAGGCCTT
                                                                          610
GTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATG-CTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG
GTGTGTCAAGTTCCAG-CCTCTGTGGCCACCCCCAGGAATGGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG
                                                                          686
GAGGAAGCCTTCCTGGGCAT-CACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCT
GAGGAAGCCTTCCTGGGCATGC-CTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCT
                                                                          763
ACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAG
ACACAAACTGGAACGAGGGTGAACCCAACAATGCTGGTTCTGATGAACATTGTGTATTGCTACTGAAAAATGGCCAG
                                                                          828
TGGAATGACGTCCCCT-GCT-CCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCTGAAGGGTCAT----------
TGGAATGAC-TCCCCTTGCTTCCACCTCCCATCTGCC-GTCTGTGAGTTCCCTATCTGAAGGGTC-TGTGAGTTCCC
                                                                          886
-------------ATCACTCA----GGCCCTCCTTGTCTTTTTACTGCA-ACCC-ACAGGCCCACAGTATGCTTGAA
TATCTGAAGGGTCAT-A-TCACTCAGGCCCTCCTTGTCTTTTTACT-C-TACCACA---GCCCAC-GTATGCTTGAA
                                                                          963
AAGATAATGATTATATCAATTTCCTCATATCCAGCATTGTTCCTTTTGTGGGCAATCACTAAAAATGATCACTAACA
A-GATAATGAT------------------------------------------------------------------
                                                                         1040
GCACCAACAAAGCAATAATAGTAGTAGTAGTAGTTAGCAGCAGCAGTAGTAGTCATGCTAATTATATAATATTTTTA
-----------------------------------------------------------------------------
                                                                         1117
ATATATACTATGAGGCCCTATCTTTTGCATCCTACAATTAATTATCTAGTTTAATTAATCTGTAATGCTTTCGATAG
-----------------------------------------------------------------------------
                                                                         1139
TGTTAACTTGC------------------------------------------------------CAGTATGAAA-A
-----------AATATATCATTCTCAGATCAGTACCTGCAGATGAAAATATCAGATCAGTACCTGCAG-ATGAA-GA
                                                                         1214
TAAGACGGA-TTTATTTTTCCATTTACAACAAACACCTGTGCTCTGTTGAGCCTTC-CTTTCTGTTTGGGTAGAGGG
TAAGACGGCATTTATTTTTCCATTTACAACAAACACCTGTG-T--GTTGAGCCTT-ACTTTCTGTTTGGGTAGAGGG
                                                                         1290
CTCCCCTAATGACATCACCACAGTTTAATACCACAGCTTTTT-ACCAAGTTTCAGGTATTAAGAAAATCTATTTTTG
CTCCCCTAATGACATGACCACAGTTTAATACCACAGCTTTTTTACCAAGTTTCAGGTATTAAGAAAATCTATTTT-G
                                                                         1342
TAACTTTCTCTATGAACTCTGTTTTCTTTCTAATGAGATATTAAACCATGTA
TAACTTTCTCTATGAACTCTGTTTTCTTTCTAATGAGATATTAAACCATGTAAAAAAAAAAAAAAAAAA
```

**4. Comparison of the nucleotide sequences of the exons of cMBP1 (upper) and the hepatic cDNA clone [12] (lower)**

he cMBP1 sequence is numbered. The three intron/exon boundaries of cMBP1 are shown as vertical lines and are numbered. he translation initiation and termination codons are marked by horizontal lines. Gaps in the sequence are indicated by yphens. See the text for details.

n the same gene, we suggest that the sequence of our omic clone is the correct sequence because the pre- ed changes increase the overall homology of human P to rat MBP C and rat MBP A, as described above.

### tream regulatory sequences

xamination of the 843 bp of sequence ob- ed upstream of the 5′ end of the cDNA clone [12] wed the presence of several consensus sequences, ch are indicated in Fig. 3. TATAA and CAAT boxes [21] are found at −38 bp and −79 bp from the start of the cDNA clone respectively. There is a region similar to the *Drosophila* heat-shock promoter consensus sequence CTnGAAnnTTCnAG [30] at −592 bp (TAAGAAATTTCCAG). In *Drosophila* the presence of heat-shock consensus sequence in the 5′ region of the heat-shock genes is necessary for heat-induced chromosome 'puffing' and the co-ordinate increase in transcription of the heat-shock gene products [30]. The heat-shock consensus element in the 5′ region

```
                                                               20
MBP  MetSerLeuPheProSerLeuProLeuLeuLeuSerMetValAlaAlaSerTyrSer -  -  -
cDNA       Cys   IleThr -   Ser                                -  -  -
MBPC             Thr    - Phe         CysValLeuThr -   Val   Ala -  -  -
MBPA    Leu   Leu     -    LeuVal         CysValVal    - Val   Ser   GlySerGlu
                                                                        40
MBP  -  -  - GLUThrValThrCysGluAspAlaGlnLysThrCysProAlaValIleAlaCysSerSer
cDNA -  -  -
MBPC -  -  -         Leu    -   Gly      SerSer        -
MBPA ThrCysGlu       LeuLys -  -  -  -  -  -  -     Ser -             -  -
                                                                        63
MBP  ProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLysGlyGluLysGlyGluProGly
cDNA                                                 Arg      GlyThr
MBPC       Leu                       His      Ala
MBPA -  -  -  -  -  -  -  -   Arg             Pro
                                                                        86
MBP  GlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProProGlyAsnProGlyProSerGly
cDNA
MBPC                                    Val      Ala   Pro      AsnPro
MBPA                                                   SerVal   AlaPro
                                                                       107
MBP  SerProGlyProLysGlyGlnLysGlyAspProGlyLysSer -  - ProAspGlyAspSerSerLeu
cDNA                                            -  -                   Pro
MBPC    Lys   AlaThr   Pro         Arg          -  - ValGluPhe   ThrThrAsn
MBPA    Gln                        Arg   Asp   ArgAlaIleGluValLysLeuAlaAsn
                                                                       130
MBP  AlaAlaSerGluArgLysAlaLeuGlnThrGluMetAlaArgIleLysLysTrpLeuThrPheSerLeu
cDNA GlyCysLeuArgLys   SerSerAlaAsnArgAsnGlyThrTyrGln
MBPC IleAspLeu   IleAla    ArgSer  LeuArgAlaMetArg   TrpValLeuLeu   Met
MBPA MetGluAla   IleAsnThr   LysSerLysLeuGluLeuThrAsn   LeuHisAla      Met
                                                                       153
MBP  GlyLysGlnValGlyAsnLysPhePheLeuThrAsnGlyGluIleMetThrPheGluLysValLysAla
cDNA
MBPC SerGluAsn      Lys   Tyr   MetSerSerValArgArg   ProLeuAsnArgAla
MBPA       LysSer   Lys         Val     His   Arg   Pro   Ser
                                                                       176
MBP  LeuCysValLysPheGlnAlaSerValAlaThrProArgAsnAlaAlaGluAsnGlyAlaIleGlnAsn
cDNA                     ProLeuTrpProPro   GlyMet
MBPC       SerGluLeu   GlyThr                    Glu      Arg
MBPA       SerGluLeuArgGlyThr      Ile           Glu      Lys         Glu
                                                                       199
MBP  LeuIleLysGluGluAlaPheLeuGlyIleThrAspGluLysThrGluGlyGlnPheValAspLeuThr
cDNA                             MetPro
MBPC ValAla   AspVal                      GlnArg      AsnVal   Glu
MBPA ValAla   ThrSer                         Val            MetTyrVal
                                                                       222
MBP  GlyAsnArgLeuThrTyrThrAsnTrpAsnGluGlyGluProAsnAsnAlaGlySerAspGluAspCys
cDNA                                                                  His
MBPC         ValArg                                  Val      Gly   Asn
MBPA    Gly         Ser      LysLysAsp          AspHis      Gly   Asp
                                                                       245
MBP  ValLeuLeuLeuLysAsnGlyGlnTrpAsnAspValProCysSerThrSerHisLeuAlaValCysGlu
cDNA                                  Ser      PheHisLeuProSer
MBPC    Val      Thr      Lys                     Asp   Phe   Val
MBPA    ThrIleValAsp      Leu         IleSer   GlnAla      Thr
           248
MBP  PheProIle
cDNA
MBPC    SerAsp
MBPA       Ala
```

**Fig. 5. Comparison of the amino acid sequences predicted from cMBP1, the human hepatic MBP cDNA clone [12] and the rat hepatic cDNA clones for MBP C and MBP A [11]**

The amino acid sequence predicted from cMBP1 is numbered. Residues identical with those predicted from cMBP1 are indicated by clear spaces. Gaps in the sequence, introduced to optimize alignment, are indicated by hyphens. The intron/exon boundaries of MBP1 are shown by vertical lines. The *N*-terminal residue (Glu) of MBP1 is in upper case. Invariant residues in the carbohydrate-binding domain are in **bold** type and underlined. See the text for details.

of the MBP gene may have an analogous function and lead to enhancement of MBP transcription during the acute-phase response. It is noteworthy that heat-shock consensus sequences are present in the 5′ flanking region of the human C-reactive-protein gene, whose product is also an acute-phase protein [31].

There are three regions of the 5′ flanking sequence of the MBP gene that are similar to consensus sequences for glucocorticoid-responsive elements. At −245 bp is a sequence (GGTCCCATTTGTTCT) similar to the glucocorticoid-responsive element GGTACAnnnTGTTCT [32–34]. At −656 bp is a sequence (AGAACAT) identical

with another glucocorticoid-responsive element [35], and at −736 bp is a sequence (AGAACAAATG) similar to the glucocorticoid-responsive element AGAACAGATG [36]. Increased release of glucocorticoid hormones by the adrenals is an important part of the response to stress and results in elevated serum concentrations, principally of cortisol. Thus the glucocorticoid-responsive elements in the promoter region of the MBP gene may also serve to increase MBP expression during the acute-phase response.

The region from −204 bp to −184 bp has 90% homology with a part of the promoter region of the cloned serum amyloid A gene (*SAAg9*) [37], whose product is another acute-phase protein. The expression of the *SAAg9* gene is increased by interleukin 1 and tumour necrosis factor $\alpha$ [38]. The 5′ region of the *SAAg9* gene, which contains the elements responsive to interleukin 1, also contains the MBP-like 5′ sequence, although it is not known if this is the sequence that is responsive to interleukin 1 (P. Woo, personal communication). The presence of a highly homologous 20 bp sequence in the promoter regions of MBP and serum amyloid A protein suggests that MBP synthesis may also be induced by cytokines and is again consistent with MBP being an acute-phase protein. Regulatory sequences have not yet been identified in the rat MBP A gene [16].

Serum amyloid A protein and C-reactive protein are two major acute-phase proteins, synthesized by the liver, which increase in concentration up to 1000-fold or more in response to tissue injury or inflammation. The details of the regulation of expression of these genes is unclear, but their 5′ regions contain heat-shock consensus elements [31] and sequences responsive to cytokines [38]. The report by Ezekowitz *et al.* [12] that hepatic MBP mRNA concentrations were low in a normal liver but were greatly increased in a human liver that had been exposed to acute stress first suggested that MBP might also be an acute-phase protein. The findings in the 5′ region of the human MBP gene of a heat-shock consensus sequence, three glucocorticoid-responsive elements and a sequence with a high degree of homology to a similar sequence in the serum-amyloid-A-protein gene lends further support to the view that MBP is an acute-phase protein. Further work is required to establish which factors enhance MBP gene expression and which sequences mediate this effect.

These data also provide insight into the possible biological role of MBP. The signal peptide and collagen-like domain of MBP as well as the evidence that it is an acute-phase protein all indicate that MBP is a hepatic secretory protein. It seems unlikely that this MBP functions as an intracellular transport protein [39]. As an acute-phase protein like serum amyloid A protein or C-reactive protein, serum MBP may be a part of the host defences acting as a carbohydrate-specific immune system. Potential ligands include bacteria, viruses and yeasts that bear mannose-rich oligosaccharides. The binding of MBP to an organism may then result in its destruction by MBP-mediated activation of the complement cascade [10].

We thank Dr. A. Sano for the MBP monoclonal antibody, Dr. Dimitris Kioussis for the cosmid library, Dr. K. Drickamer for amino acid sequencing, Dr. M. Waterfield for oligonucleotide synthesis and Dr. J. Scott, Dr. A. Ezekowitz, Dr. P. Woo and Dr. T. Lund for advice. This work was supported by the Wellcome Trust.

## REFERENCES

1. Townsend, R. & Stahl, P. (1981) Biochem. J. **194**, 209–214
2. Mizuno, Y., Kozutsumi, Y., Kawasaki, Y. & Yamashina, I. (1981) J. Biol. Chem. **256**, 4247–4252
3. Maynard, Y. & Baenziger, J. U. (1982) J. Biol. Chem. **257**, 3788–3794
4. Kawasaki, T., Etoh, R. & Yamashina, I. (1978) Biochem. Biophys. Res. Commun. **81**, 1018–1024
5. Kozutsumi, Y., Kawasaki, T. & Yamashina, I. (1980) Biochem. Biophys. Res. Commun. **95**, 658–664
6. Kawasaki, N., Kawasaki, T. & Yamashina, I. (1983) J. Biochem. (Tokyo) **94**, 937–947
7. Summerfield, J. A. & Taylor, M. E. (1986) Biochim. Biophys. Acta **883**, 197–206
8. Taylor, M. E. & Summerfield, J. A. (1987) Biochim. Biophys. Acta **915**, 60–67
9. Wild, J., Robinson, D. & Winchester, B. (1983) Biochem. J. **210**, 167–174
10. Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. & Yamashina, I. (1987) J. Biol. Chem. **262**, 7451–7454
11. Drickamer, K., Dordal, M. S. & Reynolds, L. (1986) J. Biol. Chem. **261**, 6878–6887
12. Ezekowitz, A. B., Day, L. E. & Herman, G. A. (1988) J. Exp. Med. **167**, 1034–1046
13. Drickamer, K. (1988) J. Biol. Chem. **263**, 9557–9560
14. Haagsman, H. P., Haegood, S., Sargent, T., Buckley, D., White, R. T., Drickamer, K. & Benson, B. J. (1987) J. Biol. Chem. **262**, 13877–13880
15. Halberg, D. F., Proulx, G., Doege, K., Yamada, Y. & Drickamer (1988) J. Biol. Chem. **263**, 9486–9490
16. Drickamer, K. & McCreary, V. (1987) J. Biol. Chem. **262**, 2582–2589
17. Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. & Playfair, J. H. L. (1987) EMBO J. **6**, 355–361
18. Grunstein, M. & Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 3961–3965
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
21. Breathnach, R. & Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349–383
22. Yamada, Y., Liau, G., Mudryj, M., Obici, S. & de Crombrugghe, G. (1984) Nature (London) **310**, 333–337
23. Lozano, G., Ninomya, Y., Thompson, H. & Olsen, B. R. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4050–4054
24. Kurkinem, M., Bernard, M. P., Barlow, D. P. & Chow, L. T. (1985) Nature (London) **317**, 177–179
25. Sakurai, Y., Sullivan, M. & Yamada, Y. (1986) J. Biol. Chem. **261**, 6654–6657
26. Leung, J. O., Holland, E. C. & Drickamer, K. (1985) J. Biol. Chem. **260**, 12523–12527
27. von Heigne, G. (1982) J. Mol. Biol. **159**, 537–541
28. von Heigne, G. (1983) Eur. J. Biochem. **133**, 17–21
29. Drickamer, K. (1987) Kidney Int. **32**, S167–S180
30. Simon, J. A., Sutton, C. A., Lobell, R. B., Glaser, R. L. & Lis, J. T. (1985) Cell **40**, 805–817
31. Woo, P., Korenberg, J. R. & Whitehead, A. S. (1985) J. Biol. Chem. **260**, 13384–13388

32. Schutz, G. (1988) Biol. Chem. Hoppe-Seyler **369**, 77–86
33. Martinez, E., Francoise, G. & Wahli, W. (1987) EMBO J. **6**, 3719–3727
34. Evans, R. M. (1988) Science **240**, 889–895
35. Payvar, F., De Franco, D., Firestone, G. L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J.-A. & Yamamoto, K. R. (1983) Cell **35**, 381–392
36. Speck, N. A. & Baltimore, D. (1987) Mol. Cell. Biol. **7**, 1101–1110
37. Edbrook, M. R., Burt, D. W., Cheshire, J. K. & Woo, P. (1989) Mol. Cell. Biol. **9**, 1908–1916
38. Woo, P., Sipe, J., Dinarello, C. A. & Colten, H. R. (1987) J. Biol. Chem. **262**, 15790–15795
39. Mori, K., Kawasaki, T. & Yamashina, I. (1988) Arch. Biochem. Biophys. **264**, 647–656

Received 19 January 1989/3 April 1989; accepted 27 April 1989

# NCBI Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.6 [Apr-09-2003]**

Match: 1 Mismatch: -2 gap open: 5 gap extension: 2
x_dropoff: 50 expect: 10000.0 wordsize: 11 Filter Align

---

**Sequence 1** lcl|seq_1 **Length** 1110 — TAYLOR MBP cDNA

**Sequence 2** lcl|seq_2 **Length** 1595 — SEQ ID NO: 1 FROM OUR APPL.

**No significant similarity was found**

# Blast 2 Sequences results

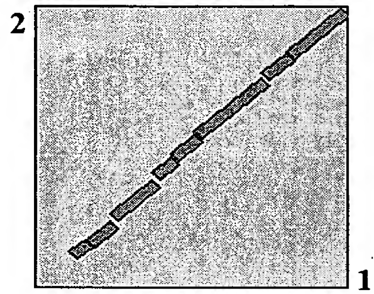
PubMed Entrez BLAST OMIM Taxonomy Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.6 [Apr-09-2003]**

Matrix BLOSUM62 gap open: 11 gap extension: 1
x_dropoff: 50 expect: 10000.( wordsize: 3 Filter Align

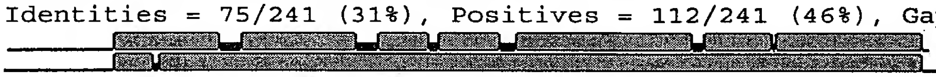
**Sequence 1** lcl|seq_1 **Length** 248 (1 .. 248) — TAYLOR MBP AA

**Sequence 2** lcl|seq_2 **Length** 277 (1 .. 277) — SEQ ID NO: 2 FROM OUR APPL.




NOTE:The statistics (bitscore and expect value) is calculated based on the size of nr database

```
Score =  107 bits (268), Expect = 2e-22
Identities = 75/241 (31%), Positives = 112/241 (46%), Gaps = 29/241 (12%)
```



```
Query: 33  PAVIACSSPGINGFPGKDGRDGTKGEKGEPGQ-------GLRGLQGPPGKLGPPGNPGPS 85
           P    C++  I+  PG  G DG KG+ GE G+        G +G++G  G +G  GN G +
Sbjct: 34  PTAEVCATHTIS--PGPKGDDGEKGDPGEEGKHGKVGRMGPKGIKGELGDMGDRGNIGKT 91

Query: 86  GSPGPKGQKGDPG-------KSPDGDSSLAASERK---ALQTEMARIKKWLTFSLG---- 131
           G  G KG KG+ G        K  G          RK    L    +AR+K  + F
Sbjct: 92  GPIGKKGDKGEKGLLGIPGEKGKAGTVCDCGRYRKFVGQLDISIARLKTSMKFVKNVIAG 151

Query: 132 -KQVGNKFFLTNGEIMTFEKVKALCVKFQASVATPRNAAENGAIQNLIKEEAF----LGI 186
            ++    KF+      E    + +   C        +A P++ A N   I + + +   F     +G+
Sbjct: 152 IRETEEKFYYIVQEEKNYRESLTHCRIRGGMLAMPKDEAANTLIADYVAKSGFFRVFIGV 211

Query: 187 TDEKTEGQFVDLTGNRL-TYTNWNEGEPNNAGSDEDCVLLLKNGQWNDVPCSTSHLAVCE 245
            D + EGQ++        L  Y+NWNEGEP++     EDCV +L +G+WND  C  +     VCE
Sbjct: 212 NDLEREGQYMFTDNTPLQNYSNWNEGEPSDPYGHEDCVEMLSSGRWNDTECHLTMYFVCE 271

Query: 246 F 246
           F
Sbjct: 272 F 272
```

```
CPU time:     0.01 user secs.      0.00 sys. secs       0.01 total secs.

Lambda     K      H
   0.312    0.134    0.405

Gapped
```

```
Lambda     K       H
   0.267   0.0410    0.140

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 939
Number of Sequences: 0
Number of extensions: 91
Number of successful extensions: 30
Number of sequences better than 10000.0: 1
Number of HSP's better than 10000.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 1
length of query: 248
length of database: 497,145,297
effective HSP length: 124
effective length of query: 124
effective length of database: 497,145,173
effective search space: 61646001452
effective search space used: 61646001452
T: 9
A: 40
X1: 16 ( 7.2 bits)
X2: 129 (49.7 bits)
X3: 129 (49.7 bits)
S1: 42 (21.8 bits)
S2: 47 (22.7 bits)
```

PATENT
Attorney Docket No. 19036/36615

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | |
|---|---|
| Applicant: Nobutaka Wakamiya<br>Serial No.: 09/600,932<br>Filed: September 8, 2000<br>For: Novel Collectin<br>Group Art Unit: 1645<br>Examiner: R. Mitra | **CERTIFICATE OF MAILING**<br><br>I hereby certify that this paper is being deposited with the U.S. Postal Service as First Class Mail in an envelope, postage prepaid, addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on November 24 2003.<br><br>Mark H. Hopkins, Ph.D. |

**DECLARATION OF NOBUTAKA WAKAMIYA**
**PURSUANT TO 37 CFR § 1.132**

I, Nobutaka Wakamiya, a citizen of the Japan, residing at 1-4, Toko-Gojo 10-chome, Asahikawa-shi, Hokkaido 078-8345, Japan, hereby declare that:

1. I am the sole inventor of the invention disclosed and claimed in the above-identified patent application, U.S. Patent Application Serial No. 09/600,932 ("the patent application"), originally filed on September 8, 2002, as a §371 of PCT/JP98/03328. As such, I am thoroughly familiar with the patent application as originally filed and amended.

2. My qualifications and technical experience are set out in my *curriculum vitae*, a copy of which is attached as Appendix A.

3. I have read and understand the official action from the U.S. Patent and Trademark Office (the "Patent Office") dated June 24, 2003 (the "Office action"), which was issued in connection with U.S. Patent Application Serial No. 09/600,932. I also have reviewed and understand the patents and publications cited by the examiner in the office action. I make this declaration to provide information known to me that may be relevant to various claim rejections in the office action.

4. Various claims stand rejected over a GenBank Accession Number AB002631 singly or in combination with the Kawai et al. publication for reasons set forth in the office action. The examiner stated that "While the reference would appear to be 24 Jan 1999, the record indicates a submission of 4 Apr 1997 which, absent factual evidence to the contrary, predates the instant application filing and the foreign priority date. The reference discloses a polynucleotide that encodes a protein of SEQ ID NO:2 and is, absent factual evidence to the contrary, a polynucleotide with the sequence of SEQ ID NO:1. Thus, claims 1, 2, and 6 are anticipated."

5. As indicated by the Examiner, the sequence data that appeared in the GenBank reference was actually submitted with GenBank on April 4, 1997 by the group of the present inventor. At the depositing of said sequence data, the present inventor simultaneously requested that GenBank not publish said sequence data until the patent application thereon was duly filed. Accordingly, the GenBank reference was published and available to the public until June 24, 1999, as is indicated by GenBank. This was one year after the filing date of the International PCT patent application PCT/JP98/03328, which was July 24, 1998.

6. For these reasons, the GenBank reference cited by the examiner does not anticipate the claims of the patent application under 35 U.S.C. 102(b). Likewise, the GenBank reference cited by the examiner cannot properly be used as an obviousness reference under 35 U.S.C. 103.

7. I further declare that all statements made herein of my own knowledge are true, that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. § 1001 and that such willful false statements may jeopardize the validity of the application or document or any patent which may issue thereon.

Nobutaka Wakamiya

Nobutakâ WAKAMIYA

Date: November 14, 2003

U.S. Patent Application No. 09/600,932
(U.S. Phase of PCT/JP/98/03328)

# CURRICULUM VITAE OF NOBUTAKA WAKAMIYA, M.D.

## [Academic and Professional Background]

**March 1980** Graduated from Hirosaki University, Faculty of Medicine, School of Medicine.

**July 1980** Joined to Osaka Prefectural Hospital as Resident Pediatrician.

**April 1982** Admitted to Doctor`s Course in Department of Pathology (Virus III) at Graduate School of Medicine, Osaka University.

**March 1986** Graduated from the Doctor`s Course above. Earned a degree of Doctor of Medicine.

**June 1986** Enrolled in Dana-Farber Cancer Institute, Harvard University as a Research Fellow.

**Jan. 1988** Joined to Research Institute for Microbial Diseases, Osaka University as an Assistant Professor.

**Nov. 2000** Joined Asahikawa Medical College as a Professor and chairman (the present position).

**[Specialized Field]**
Biochemistry, Immunology and Microbiology

**[Membership of Societies Belonged]**
The Japanese Biochemical Society
The Japanese Society for Virology
The Japanese Society for Immunology
American Society for Microbiology (ASM)
American Association for Cancer Research (AACR)

**[The Present Research Theme]**
Biological Functions and Roles of Animal Lectin (Collectin) possessing Collagen Motifs.

**[Recent Scientific Japanese Publications on Collectin]**


(1) Wakamiya N. and Suzuki Y., `Collectin Families to be acted as Bio-Defensive Lectin`, PROTEIN, NUCLEIC ACID AND ENZYME, Vol.45, No. 5, pp.655-663 (2000).

(2) Wakamiya N. and Suzuki Y., `Novel Hemangioendotheliocyte Scavenger Receptor CL-P1`, SEIKAGAKU, Vol.73, No. 3, pp.205-208 (2001).


**[Recent Scientific English Publications on Collectin]**


1. Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., Wakamiya, N.: Molecular cloning of a novel collectin from liver (CL-L1). J. Biol. Chem. 274(19): 13681-13689, 1999.
2. Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T., Itabe, H., Suzutani, T., Ogasawara, M., Yoshida, I., Wakamiya, N.: The membrane-type collectin CL-P1 is a scavenger receptor on vascular endothelial cells. J. Biol. Chem. 276(47): 44222-44228, 2001.


--**-----**-----**-----**-----**-----**-----**-----**-----**--

**Terre Robertson**



**From:** ddbjupdt@ddbj.nig.ac.jp
**Sent:** Friday, October 03, 2003 3:27 AM
**To:** Terre Robertson
**Cc:** ytateno@genes.nig.ac.jp; hsugawar@genes.nig.ac.jp; ddbjupdt@ddbj.nig.ac.jp
**Subject:** RE: Submission Publication Date Verification

Dear Dr. Terre Robertson

The sequence data with accession number AB002631 were released from the DNA Data Bank of Japan (DDBJ) on June 24 1999 in order to make them public.

DDBJ is in collaboration with the EMBL Nucleotide Sequence Database in Europe and GenBank in USA to form and function as the International Nucleotide Sequence Databases.

We take no responsibility for the priority and property issues for the submitted data. We simply inform you of the releasing date on request.

We appreciate your understanding and cooperation.

Sincerely yours,

Yoshio Tateno, Ph.D.
The Center for Information Biology and DNA Data Bank of Japan
National Institute of Genetics

---------------------------------------------
We are not using "GI" and division is "HUM".
As we told you before, please check AB002631 by using our retrival tool
getentry: http://getentry.ddbj.nig.ac.jp/getstart-e.html

---------------------------------------------

>Date: Tue, 30 Sep 2003 08:54:01 -0500
>From: "Terre Robertson" <TRobertson@marshallip.com>

>Dear Sir or Madam: You are correct that the acession no. is AB002631.
>Version is AB002631.1 GI:5162874. Locus is AB002631 1594 bp mRNA
>linear PRI-24 Jun 1999. Authors are Ohtani, Suzuki, Eda, Kawai, Kase,
>Yamazaki, Keshi, sakai, Fukuoh, Sakamoto and Wakamiya. The Direct
>Submission was made on 4 April 1997. Keywords are collectin 34.
>
>Medline no. is 99240768
>PubMed No. ia 10224141
>
>Thank you so very much for your kind assistance with this matter. Your
>help is deeply appreciated.
>
>-----Original Message-----
>From: ddbjupdt@ddbj.nig.ac.jp [mailto:ddbjupdt@ddbj.nig.ac.jp]
>Sent: Monday, September 29, 2003 8:31 PM
>To: Terre Robertson
>Cc: romiti@ncbi.nlm.nih.gov; ddbjupdt@ddbj.nig.ac.jp
>Subject: Re: Submission Publication Date Verification
>
>
>Dear Dr. Terre Robertson
>
>DNA Data Bank of Japan (DDBJ) has received your message at its update
>email address by way of GenBank again.

>As we explained in our previous mail, the accession number
>of DDBJ/EMBL/GenBank is consisted of
>* 1 alphabet character and 5 digits(ex. D12345) *** or ****
>* 2 alphabet characters and 6 digits(ex. AB123456).
>
>We think your request is about AB002631.
>Please check to be sure that it is AB002631 by using our retrival tool
>getentry: http://getentry.ddbj.nig.ac.jp/getstart-e.html
>
>After having your confirmation, we will respond you about the data
>as far as possbile. Please reply directly to us:ddbjupdt@ddbj.nig.ac.jp.
>Thank you for your cooperation.
>
>Sincerely,
>DDBJ update
>
>>Date: Mon, 22 Sep 2003 15:26:02 +0900 (JST)
>>From: ddbjupdt@ddbj.nig.ac.jp
>>Subject: Re: group: submission publication date verification
>>To: TRobertson@marshallip.com
>>Cc: romiti@ncbi.nlm.nih.gov, ddbjupdt@ddbj.nig.ac.jp
>
>>The accession number of DDBJ/EMBL/GenBank is consisted of 1 alphabet
>>character
>>and 5 digits(ex. D12345) or 2 alphabet characters and 6 digits(ex.
>AB123456).
>>But your request is about "AB0022631". Please check and let us know
>the acc
>>again. We will respond you about the data as far as possbile.
>
>>Date: Mon, 29 Sep 2003 16:35:09 -0400 (EDT)
>>From: Monica Romiti <romiti@ncbi.nlm.nih.gov>
>>Subject: Submission Publication Date Verification
>>To: ddbjupdt@ddbj.nig.ac.jp
>
>>------------- Begin Forwarded Message -------------
>
>>Subject: Submission Publication Date Verification
>>Date: Mon, 29 Sep 2003 15:25:39 -0500
>>From: "Terre Robertson" <TRobertson@marshallip.com>
>>To: <romiti@ncbi.nlm.nih.gov>
>
>
>>Dear Monica: Thank you for your email responses while I was out of
>>town last week. On the printout from NCBU Sequence Viewer, there is no
>
>>D number. I received some email from the people at DDBJ stating that
>>there would be a 6 digit number beginning with D--unfortunately in my
>>reviewing all my emails, that one got deleted. However, if you can
>>recontact them I would appreciate it. The acession no. is AB002631.
>>Version is AB002631.1 GI:5162874. Locus is AB002631 1594 bp mRNA
>>linear PRI-24 Jun 1999. Authors are Ohtani, Suzuki, Eda, Kawai, Kase,
>
>>Yamazaki, Keshi, sakai, Fukuoh, Sakamoto and Wakamiya. The Direct
>>Submission was made on 4 April 1997. Keywords are collectin 34.
>>
>>Medline no. is 99240768
>>PubMed No. ia 10224141
>>
>>This is all the information contained on the printout so I do not know
>>how to find the "D" number mentioned by the Japanese associates. Your
>>help will be deeply appreciated. If you need me to fax the papers to
>>you, please provide me with a fax number. Thanks so much for all your
>>help.
>>
>>
>>The material in this transmission contains confidential information

>>intended only for the addressee. If you are not the addressee, any
>>disclosure or use of this information by you is strictly prohibited.
>>If you have received this transmission in error, please delete it,
>>destroy all copies, and notify Marshall, Gerstein & Borun LLP by
>>telephone (312) 474-6300. Thank you.
>>
>>
>>------------- End Forwarded Message -------------
>>
>>
>>Best regards,
>>
>>Monica L. Romiti
>>NCBI User Services
>>
>
>
>
>The material in this transmission contains confidential information
>intended only for the addressee. If you are not the addressee, any
>disclosure or use of this information by you is strictly prohibited.
>If you have received this transmission in error, please delete it,
>destroy all copies, and notify Marshall, Gerstein & Borun LLP by
>telephone (312) 474-6300. Thank you.

# Blast 2 Sequences results

PubMed Entrez BLAST OMIM Taxonomy Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.6 [Apr-09-2003]**

Match: 1 Mismatch: -2 gap open: 5 gap extension: 2
x_dropoff: 50 expect: 10000.0 wordsize: 11 Filter Align

**Sequence 1** lcl|seq_1 **Length** 1409 – BOVINE MBP (KAWAII ET AL.)

**Sequence 2** lcl|seq_2 **Length** 1595 – SEQ ID NO: 1 OF 10036/36615

**No significant similarity was found**

# Blast 2 Sequences results

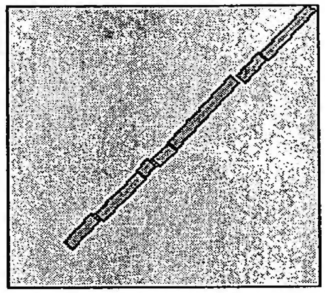
PubMed Entrez BLAST OMIM Taxonomy Structure

**BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.6 [Apr-09-2003]**

Matrix BLOSUM62 gap open: 11 gap extension: 1
x_dropoff: 50 expect: 10000.0 wordsize: 3 Filter Align

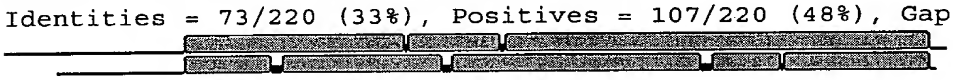
---

**Sequence 1** lcl|seq_1 **Length** 277 (1 .. 277) — SEQ ID NO: 2 OF 19036/36615
**Sequence 2** lcl|seq_2 **Length** 249 (1 .. 249) — BOVINE MBP (KAWAII ET AL.)

2

1

NOTE:The statistics (bitscore and expect value) is calculated based on the size of nr database

```
Score =  108 bits (270), Expect = 1e-22
Identities = 73/220 (33%), Positives = 107/220 (48%), Gaps = 15/220 (6%)
```



```
Query: 56   GDPGEEGKHGKVGRMGPKGIKGELGDMGDRGNIGKTGPIGKKGDKGEKGLLGIPGEKGKA 115
            G PG  G  GK GR G KG KGE G    +G  G  GP GK G +G  G+ GIPG  G+
Sbjct: 40   GPPGINGIPGKDGRDGAKGEKGEPG----QGLRGSQGPPGKMGPQGTPGIPGIPGPIGQK 95

Query: 116  GTVCD-CGRYRKFVGQLDISIARLKTSMKFVKN--VIAGIRETEEKFYYIVQEEKNYRES 172
            G   +  G Y +         A L++ +  +KN  + +  +   +K ++   ++  + E
Sbjct: 96   GDPGENMGDYIRLATS---ERATLQSELNQIKNWLIFSLGKRVGKKAFFTNGKKMPFNEV 152

Query: 173  LTHCRIRGGMLAMPKDEAANTLIADYVAKSGFFRVFIGVNDLEREGQYMFTDNTPLQNYS 232
             T C    G +A P +   N  + D V +  F    +G+ D E EG+++      +  Y
Sbjct: 153  KTLCAQFQGRVATPMNAEENRALKDLVTEEAF----LGITDQETEGKFVDLTGKGV-TYQ 207

Query: 233  NWNEGEPSDPYGHEDCVEMLSSGRWNDTECHLTMYFVCEF 272
            NWN+GEP++    E CV +LS G WND  C  +   VCEF
Sbjct: 208  NWNDGEPNNASPGEHCVTLLSDGTWNDIACSASFLTVCEF 247

CPU time:     0.01 user secs.      0.00 sys. secs         0.01 total secs.

Lambda     K      H
   0.314    0.136    0.417

Gapped
Lambda     K      H
   0.267   0.0410    0.140
```

```
Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 973
Number of Sequences: 0
Number of extensions: 109
Number of successful extensions: 32
Number of sequences better than 10000.0: 1
Number of HSP's better than 10000.0 without gapping: 1
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 5
length of query: 249
length of database: 498,883,957
effective HSP length: 124
effective length of query: 125
effective length of database: 498,883,833
effective search space: 62360479125
effective search space used: 62360479125
T: 9
A: 40
X1: 16 ( 7.3 bits)
X2: 129 (49.7 bits)
X3: 129 (49.7 bits)
S1: 42 (21.9 bits)
S2: 47 (22.7 bits)
```